



# IJRASET

International Journal For Research in  
Applied Science and Engineering Technology



---

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

---

**Volume:** 12    **Issue:** V    **Month of publication:** May 2024

**DOI:** <https://doi.org/10.22214/ijraset.2024.62137>

[www.ijraset.com](http://www.ijraset.com)

Call:  08813907089

E-mail ID: [ijraset@gmail.com](mailto:ijraset@gmail.com)

# Revolutionizing Disease Prediction with Deep Learning and Predictive Analysis

Asst. Prof. Mr. Gaurav Singhania<sup>1</sup>, Nikhil Maurya<sup>2</sup>, Parth Sharma<sup>3</sup>, Ronit Bhardwaj<sup>4</sup>

Department of Computer Science and Engineering Galgotias College of Engineering and Technology Greater Noida, Uttar Pradesh, India

**Abstract:** *The integration of deep learning and predictive analysis has emerged as a powerful tool in the healthcare improvement effort, revolutionizing the process of illness prediction. The purpose of this project is to develop a comprehensive framework that can precisely predict and identify a range of diseases at an early stage using these technologies. Deep learning algorithms are used to uncover complicated patterns and connections that may point to the onset of a disease by evaluating massive datasets that include patient demographics, medical history, genetic data, and environmental factors. The predictive capabilities of these algorithms enable healthcare practitioners to more precisely identify potential health issues, enabling preventive intervention and customized treatment regimens. Moreover, putting this idea into reality promises to improve patient outcomes, substantially reduce healthcare costs, and ultimately alter the landscape of preventative medicine.*

**Keywords:** *Deep learning; convolutional neural networks; Machine-learning; Recurrent neural networks.*

## I. INTRODUCTION

Thanks to the combination of deep learning, artificial intelligence (AI), and predictive analytics, there has been a paradigm shift in healthcare recently, particularly in the field of disease prediction and prevention. Thanks to its revolutionary synergy, early identification, more effective treatment plans, and tailored treatments, healthcare might be approached in a completely new way. Conventional healthcare models usually focus on reactive treatment, or treating diseases after they have already manifested symptoms. But given the advent of cutting-edge technology and the exponential growth of medical data, there is a compelling argument to shift to a proactive paradigm that primarily focuses on predictive analysis.

The goal of this project is to create a strong framework that will transform disease prediction by utilizing the power of deep learning and predictive analytics. We can train advanced AI models to identify the subtle patterns, and the correlation indicative of illness development by utilizing the large datasets that cover a variety of patient information, such as demographics, medical history, genetic markers, lifestyle factors, and environmental effects. By utilizing deep learning methods, such as gradient boosting machines (GBMs), recurrent neural networks (RNNs), and convolutional neural networks (CNNs), we can extract hidden patterns from these intricate datasets.

These kinds of innovations have far-reaching effects. By accurately forecasting the likelihood of illness development well in advance, healthcare providers can take proactive measures to lower risks and improve patient outcomes, such as early intervention, lifestyle modifications, and individualized treatment regimens. Moreover, integrating predictive analytics into healthcare systems has the potential to significantly reduce healthcare costs by proactively addressing health conditions before they progress into more catastrophic illnesses. Large datasets that cover a variety of patient information, such as demographics, medical history, genetic markers, lifestyle factors, and environmental effects. By utilizing deep learning methods, such as gradient boosting machines (GBMs), recurrent neural networks (RNNs), and convolutional neural networks (CNNs), we can extract hidden patterns from these intricate datasets. This introduction establishes the foundation for further research into the revolutionary potential of deep learning and predictive analysis, with the goal of revolutionizing disease prediction and paving the way for a future in which healthcare is more than just a means of curing illness.

The effects of this undertaking are extensive and extend well beyond the provision of medical treatment. By giving clinicians rapid, actionable insights, we may improve patient outcomes and create considerable cost and resource savings throughout the healthcare continuum. Additionally, by proactively addressing health inequities and stopping the course of the disease, we may foster the culture of resilience and well-being that transcends traditional notions of illness and therapy.

We would like to invite everyone involved in the healthcare system to work with us as we embark on our mission to revolutionize disease prediction. Our vision is to build a future where healthcare is proactive rather than reactive, enabling every individual to lead a long and healthy life.

## II. LITERATURE SURVEY

The goal of Senthil Kumar Mohan et al.'s technique, Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques, is to increase the precision of cardiovascular disease prediction by using machine learning to pinpoint key variables. To create the expectations model, several combinations of highlights and a few well-known arranging strategies are applied. By utilizing the hybrid random forest with a linear model (HRFLM) prediction model, we can get an enhanced display level with an accuracy rate of 88.7%. Additionally, they learn about a variety of data mining methodologies and expectation techniques, including KNN, LR, SVM, NN, and Vote, which have garnered notoriety recently for their capacity to detect and forecast heart disease.

Sonam Nikhar et al. developed the paper Prediction of Heart Disease Using Machine Learning Algorithms. The goal of this study is to give a thorough description of the Naive Bayes and decision tree classifiers that we employed in our analysis, especially for the heart disease prediction. It has been found through certain research that Decision Trees perform better than Bayesian classification systems when predictive data mining techniques are used to comparable datasets.

Aditi Gavhane, Isha Pandaya, Gouthami Kokkula, Prof. Kailas Devidkar (PhD), and Heart Disease Forecast The multi-layer perceptron (MLP) neural network approach was used in this paper's recommended system to train and test the dataset through machine learning. This approach involves a number of layers, including two for input and output and one or more hidden layers between the two for input and output. Every input layer node is connected to every output layer node via these hidden layers. This link has weights attached to it. A second identical input called bias with weight  $b$  is introduced to the node in order to balance the perceptron. The connections between the nodes may be feed-forward or feed-back, depending on the requirements.

The Heart Attack Prediction Using Deep Learning was developed by Abhay Kishore and colleagues. This work proposes a heart attack prediction system that estimates the patient's risk of contracting a heart-related infection using deep learning techniques, namely a recurrent neural system. The Recurrent Neural Network is a very creative characterization computation that uses the Deep Learning technique to Artificial Neural Networks. In this study, the key framework modules and the related hypothesis are covered in detail. The suggested methodology produces precise findings with low error rates by utilizing deep learning and data mining. This book offers direction and a point of reference for the creation of an alternative platform for heart attack prediction. phase of forecasting.

Methods of Machine Learning for Heart Disease According to Lakshmana Rao et al., these areas are more likely to have high blood pressure, diabetes, circulatory strain, and current smoking, among other heart disease risk factors. As a result, it is difficult to distinguish between heart diseases. Individuals' cardiac disease severity has been assessed through a range of data mining and neural system techniques. Because the idea of the CHD sickness is unclear, caution must be used when treating the ailment. If a patient is not found in time, it could have a negative impact on their heart or cause an unplanned death. From the perspective of medical research, data burrowing is also utilized to identify various forms of metabolic machine learning, which is a method by which the framework learns from earlier data tests and models without being specially tailored. Machine learning forces reasoning to rely on past data.

Dr. Santhana Krishnan J. and Dr. Geetha S., Predicting heart disease with machine learning This study predicts a male patient's likelihood of having heart illness using categorization approaches. This paper offers a thorough description of coronary heart disease, covering its causes, prevalent forms, and risk factors. The DataMining tool being used is WEKA (Waikato Environment for Knowledge Analysis), a great data mining tool for bioinformatics domains. The three WEKA interfaces that are available in this system—decision trees, artificial neural networks, and naive bayes—are used to forecast heart disease. These are the principal methods used in data mining. The main technique for prediction is decision trees, such as CART, C4.5, CHAID, J48, ID3 Algorithms, and Naive Bayes Techniques.

Forecast for Heart Illness, according to Avinash Golande et al., the use of a few data mining techniques by specialist's aids in the identification of heart sickness by doctors or authorities through the use of Effective Machine Learning Techniques. Decision trees, k-closest, and Naïve Bayes are commonly utilized as techniques. Additional unique characterization-based methods include part thickness, packing computation, neural systems, SVM (Bolster Vector Machine), sequential negligible streamlining, and straight kernel self-arranging guidance. Of course, information regarding the systems employed in the analysis is given in the next section.

V.V. Ramalingam et al. recommended the use of machine learning techniques and algorithms for the prediction of cardiac disease in a range of medical datasets in order to automate the analysis of large and complex volumes of data. Many researchers have recently begun using different machine learning techniques to help specialists in the medical field identify heart-related disorders. This paper presents a survey of many models constructed with these kinds of algorithms and approaches, and analyzes the results. Models based on supervised learning techniques such as Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Naïve Bayes, Decision Trees (DT), Random Forests (RF), and ensemble models are among the models that researchers choose.



These systems have been used to automate the examination of enormous volumes of intricate data across a variety of clinical datasets. Many researchers have recently automated the processing of large and complex data sets by using a few machine learning techniques and algorithms on a range of medical datasets. Many researchers have recently begun using different machine learning techniques to help specialists in the medical field identify heart-related disorders. This study examines a number of models that are based on these kinds of techniques and algorithms and assesses how well they perform. Researchers discover that models built using supervised learning algorithms are very popular. Examples of these models are Random Forest (RF), Decision Trees (DT), K-Nearest Neighbor (KNN), Naïve Bayes, Support Vector Machines (SVM), and ensemble models. methods to support medical professionals and the health sector in the identification and management of heart-related conditions. This paper evaluates and investigates the display of several models that rely on these computations and approaches. When it comes to directed learning computations, models like Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Naïve Bayes, Decision Trees (DT), Random Forests (RF), and group models are found to be highly well-known among academics.

Huiyan Wang has established himself in this field by releasing a computerized diagnostic model in an attempt to standardize and boost the adoption of traditional Chinese medicine (TCM) diagnosis (Huiyan Wang 2008). This is the way the system functions. To choose the symptoms, the Bayesian network topology is first understood using a database of examples that incorporates mutual information theory. The Markov blanket of the target variable is selected as the symptom set in the structure. Second, the mapping relationship between symptom sets and diagnostic results is supported by naive Bayesian classifiers.

The mission assigned to Huang Jianyong in the field was as follows (2004). This study aimed to address the multi-class classification problem by means of binary classifiers. In the binary classification tree design, a single node splits the class group into two separate subgroups. The node improves its ability to classify binary data by employing the class module technique. Partitioning can be formulated as an optimization issue, and genetic algorithms provide solutions to optimization problems. Heart murmurs were a major problem in pediatric cardiovascular disease, and data indicated that this group had a high incidence of cardiac noise (77–95%), which helped to alleviate congestion associated to congenital heart disease. Sanjay developed a reliable diagnostic test in 2005 for pediatric murmurs.

Carlos Ordonez looked at the application of association principles in heart disease prognosis. They used a simple mapping technique. Both numerical and category properties are handled by this approach. It is used to transform medical records into transaction formats. Improved methods are used to minify restricted association rules. The mapping table is prepared and the attribute value is mapped to the object. Decision trees are used in data mining because they automatically split numerical values. Split points selected in the decision tree are rarely used. Clustering is used to understand the data as a collective entity.

To extract association rules, Deepika examined the Pruning-Classification Association Rule. Apriori algorithm analysis is the basis of PCAR. In order to exclude irregular items from a group of items, PCAR covers a set of minimum frequency entries. It then identifies a frequent set of items by sorting the set of items according to how frequently they occur. Ordonez use association rules in data mining to get more accurate predictions of heart disease. The limits of association rules that mined the complete data set without independent sample verification were examined by the authors. In order to decrease the number of association rules, we devised algorithms that changed the search constraints. We then used train and test methods to validate them. They have looked at two related tasks: determining whether heart disease is present or absent.

Using feature subset selection, P. Chandra and M. Jabbar created class association rules to detect heart disease. Association rules determine the relationship between attribute values and classification to forecast classes in a patient data collection. Genetic searching and other functional selection strategies are used to identify the features that help predict heart disease.

Usha Rani proposed a method for forecasting cardiac disease using feedforward and backpropagation algorithms in artificial neural networks. We explore with neural network models that are single-layer and multi-layer. Parallel processing is employed to accelerate the learning process in each neuron across all hidden output layers.

Ensan et al. (2006) introduced the Fuzzy Clustering (FACT) technique. It says that the density of the data must determine how many clusters fit the data. However, the proposed method is insensitive to initial cluster counts, which are often changed to be less than the threshold number of clusters. Their method generates several clusters by generating new clusters to focus on the outlier discovery procedure. They conducted an empirical study and found that the recommended heuristic approach outperforms the conventional K-means computation. The SVM classification approach for DM detection in machine learning is described by Purnami et al. (2009).

Eventually, Aishwarya et al. (2013) proposed a fuzzy logic-based solution for DM analysis. In their model, they established two connections between symptoms and diseases: an occurrence relationship and a confirmability relationship.

While the occurrence relationship attests to a symptom's repeated presentation, the confirmability relationship shows how severe symptoms are in proportion to the disease's presence. Additionally, they recommended applying fuzzy logic with minimum and maximum links while analyzing DM, and they used a real-world dataset of 40 patients to determine the fuzzy link. Kumari and Chitra (2013) have also linked a few ML and data mining techniques for the prediction of DM. Their work made use of SVM. According to exploratory experiments on a genuine diabetic dataset, understandable SVMs give a promising accuracy for DM prediction; investigational results indicated an accuracy of 79.00%. It has been noted that SVM classifiers can yet become more effective by using a feature subset selection process.

Kumar et al. (2014) proposed using Genetic Algorithm (GA) to discover a correct component subclass with SVM classifier on a range of datasets in order to improve the characterization precision. The GA-based classifier seems to increase the border esteems for SVM. When the outcomes were identified using the present SVM methodology, the suggested strategy yielded an accuracy of 83.00%.

To predict individuals with diabetes, Jhaldiyil and Mishra (2014) also used Principal Component Analysis (PCA) and Support Vector Machines (SVM). The investigational findings of the study indicated that, with a classification accuracy of almost 93.66%, the previous level might be enhanced.

Vispute et al. used order processes to investigate the risk of DM in their 2015 study. This paper evaluated four machine learning techniques: naive bayes, logistic regression, artificial neural networks, and decision trees. After receiving the data, the application generates an output that compares the actual and projected figures. In the end, their investigation demonstrated that Random Forest attains a high degree of accuracy. Around the same time, Lingaraj et al. (2015) experimented with DM detection using the NB classifier and the WEKA tool. The study's data came from Indian hospitals and comprised 1865 examples of blood and urine test features. The accuracy attained, even with the authors' ten-fold validation, is just 84.89%, which is incredibly low and calls for more investigation.

Recently, Perveena et al. (2016) employed DT J48 for DM detection, subject to risk variables. They demonstrated how Adaboost outperforms DT J48 and bagging in their system. However, the absence of basis learners in the ensemble architecture causes this work to delimit in numerous performance metrics, resulting in a research vacuum that will be filled by including base learners in an ensemble of classifiers. Vashi and Mishra (2016) went into greater information regarding the types and causes of diabetes that are relevant to the diagnosis of diabetes mellitus in their survey study of DM prediction. The authors focused mostly on classification and clustering techniques for DM detection in addition to neural networks. Nevertheless, no actual data were reported in the Vashi-Mishra work.

Vrushali Balpande & Wajgi (2017) investigated classification strategies for the identification of DM utilizing the DT and NB classifier in the WEKA tool. They concluded that the NB classifier is not as effective as the DT technique. Anjali (2017) uses supervised learning techniques and neural networks (NNs) to predict the overall survivability of DM in comorbidity. The researcher proposed a PCA-based methodology to minimize the dimension of recovered features using NN as the classifier. Ninety-two.2% of the findings were precise. It is still possible to improve the work's performance by modifying the NN's settings. When employing PCA, the need for data regularization frequently adds unnecessary expense to the process.

Edla et al. suggested the creation of a diabetes decision support system (DSS) (2017). A radial basis function NN classifier (RBFNN) is used in its construction. RBFNN is a three-layer neural network: the first layer handles the model's contributions; the second layer, known as the buried layer, is made up of several non-direct RBF actuation units; and the third layer is the yield layer of the neural organization classifier. RBFNN uses Gaussian capabilities to perform enactment tasks. Using the PIMA Indian dataset, RBFNN builds a DM expectation model with 73.91% accuracy, 81.33% sensitivity, and 60% specificity. The limitation of the approach is the trade-off between accuracy and the number of concealed layers.

Cui et al. (2018) suggested a hybrid estimate model that applies PCA to the original dataset and then uses C4.5 methods to generate the classifier model. The classification accuracy of their work was only about 89.0%, but with the appropriate feature selection approaches, it can be significantly improved.

Haritha et al. (2018) discussed a different data mining technique for DM detection that blends association rule mining with classification methods. The authors used the KNN classifier to assess the classification execution strategy. The authors initially created the KNN analysis with ten characteristics, and 61.9% accuracy was attained. The accuracy was then raised to 88.5% by the authors using PSO techniques to choose six features for training.

Using search strategies like firefly and cuckoo, Rashid and Abdullah (2018) discussed selecting features from the Indian PIMA dataset for Type-1 and Type-2 DM identification. For the optimal features discovered by cuckoo search, they used a fuzzy-KNN classifier; for the DM prediction, they used a KNN classifier based on the ideal features discovered through the firefly technique.

After doing an experimental examination, they concluded that the cuckoo-Fuzzy KNN had the best classification accuracy. Moreover, the learning rate of the optimization algorithm can be raised.

Zhang et al. (2018) discussed diagnosing DM with hybrid ant colonies, GA, and NNs. The author took into account the Indian PIMA dataset, which selects features using ant colonies and GA. On the selected features, they then applied back propagation NNs. Through trials, they proved that they could use this model to achieve good accuracy. CPU time taken while hybridizing optimization techniques becomes an increasing issue that needs to be addressed in applications that are comparable. Kadhm, et al. (2018) addressed the feed forward neural network approach to the DM identification challenge. Using the Indian PIMA diabetes dataset, the author used a two-layer feed forward neural network approach. They proved via experimentation that the FFNN model is capable of above 82% accuracy.

Perveen et al. (2019) demonstrated a sophisticated SVM model for DM diagnosis. These investigators, who see diabetic mellitus (DM) as a serious worldwide health concern, found that 80% of DM complications are treatable if detected early in the disease. The proposed model has explored a number of data mining and machine learning methodologies for the DM forecast. For example, they proposed to use an extra module to the SVM model to convert the "discovery" model into an acceptable representation. This novel design offers an exact and unambiguous SVM grouping conclusion. For each and every sample scenario, there is still potential for improvement in terms of performance enhancement. On the other hand, Zhu et al. (2019) recommended utilizing PCA to describe the dimensional datasets and a tweaked version of K-Means to extract the pertinent data. Their strategy used the distinct highlight subcategory in the second stage and grouped instances in the first to determine the best classifier for DM prediction. The PCA subset and avalanche SVM were found to have contributed to the best level of 81.28% classification accuracy for the SVM, based on test results. The lack of a method for determining the number of major components will place restrictions on algorithms that employ PCA.

In order to give patients convenient medication through the usage of unique AI models, Sivakumar et al. (2020) aim to find results for the diagnosis of diabetes mellitus by analyzing the useful examples included in the data. They attained 76.3% and 75.7% precision for examples that AI classifiers, such the NB computation and RF technique, effectively ordered. Srivastava et al. (2020) created a model that combines SVM with the Fuzzy C-Means Clustering (FCM) algorithm to anticipate diabetes. The model's accuracy was 84.24%. Another advantage of this work is that the authors did not make any compromises on the computational trade-off that arises from removing undesirable data using a fuzzy C-means clustering technique.

Table 1: Literature Overview

Author	Title	Technologies Used	Advancement (Classifier - Accuracy (%))
Senthil Kumar Mohan	Predict heart diseases	Hybrid machine learning techniques	88.7
Sonam Nikhar	Prediction of Heart diseases	Machine learning algorithms, decision tree	87.6
Aishwarya	Prediction of Heart Disease	Fuzzy logic	79
Kumar, et al.	Heart Prediction	SVM Classifier	83
Jhaldial and Mishra	Predict diabetes diseases	SVM	93.66

Vispute et al.	Predict Diabetes diseases	Decision tree, ANN, LR, NB	84.89
----------------	---------------------------	----------------------------	-------

Vrushali Balpande and Wajgi	Prediction of Diabetes Disease	DT and NB Classifier	73.91
Cui et al.	Diabetes Prediction	KNN Classifier	89
Zhang, et al.	Prediction of Diabetes Disease	FFNN, SVM, DT	82
Perveen, et al.	Diabetes Prediction	SVM	81.28

### III. PROPOSED WORK

Our proposed effort develops an advanced illness prediction system for diabetes and heart disease using cutting edge techniques from deep learning, machine learning, and predictive analysis. The primary objective of this work is to create accurate and reliable prediction models that can detect high-risk individuals for diabetes and heart disease, enabling personalized preventive care and early intervention.

To do this, we will employ a multidisciplinary approach that integrates numerous datasets including clinical, demographic, genetic, and lifestyle factors associated with diabetes risk and cardiovascular health. By cleaning, normalizing, and engineering features, data preparation procedures will be applied to guarantee the quality and integrity of the input data for the building of the model.

Next, we will explore several deep learning architectures including recurrent neural networks (RNNs) and convolutional neural networks (CNNs), in addition to more traditional machine learning techniques like logistic regression and random forests. These models will be trained and validated using integrated datasets, and their effectiveness will be evaluated based on pre-established standards.

We will create models and apply predictive analysis techniques to generate prognostic indicators and risk ratings for diabetes and heart disease. Cross-validation will be used to ensure the system's resilience and generalizability over a range of population categories. The ultimate objective of the suggested endeavor is to advance predictive analytics in the healthcare industry and support the development of precision medicine approaches to illness management and prevention.

### IV. CONCLUSION

In summary, a significant advancement in the field of healthcare analytics is the illness prediction system, which seeks to enhance early diagnosis and tailored management of diabetes and heart disease. Our objective is to construct dependable predictive models that accurately identify individuals who are more prone to suffer from cardiovascular issues and acquire diabetes by utilizing state-of-the-art techniques in deep learning, machine learning, and predictive analysis.

Several datasets related to clinical, genetic, and lifestyle characteristics will be integrated to offer a comprehensive basis for the development and validation of the model. By using feature engineering and rigorous preprocessing, we will ensure that the input data is high-quality and pertinent, which will increase the generalizability and dependability of the prediction models.

All things considered, it holds great promise for improving disease management and prevention strategies, which will ultimately lead to better health outcomes and cheaper treatment costs for conditions like diabetes and heart disease. By collaborating with stakeholders and healthcare providers to translate our research findings into useful applications that benefit patients and healthcare systems alike, we seek to usher in a new era of precision medicine and individualized healthcare.

### REFERENCES

- [1] Jhaldiyal T. Mishra P. K. (2014). Analysis and Prediction of Diabetes Mellitus Using PCA, REP and SVM. International Journal of Engineering and Technical Research, 2(8), 164–166.



- [2] Sonam Nikhar, A.M. Karandikar Prediction of Heart Disease Using Machine Learning Algorithms International Journal of Advanced Engineering, Management and Science (IJAEMS) Infogain Publication,[Vol-2, Issue-6, June- 2016].I.S. Jacobs and C.P. Bean, Fine particles, thin films and exchange anisotropy, in Magnetism, vol. III, G.T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271-350.
- [3] Aditi Gavhane Heart diseases image classification based on Convolutional Neural Network. 2019 International Conference on Computational Science and Computational Intelligence.
- [4] A.Lakshmanarao, Y.Swathi, P.Sri Sai Sundareswar, Machine Learning Techniques For Heart Disease Prediction, International Journal Of Scientific & Technology Research Volume 8, Issue 11, November 2019.
- [5] Zhang, G. (2018, November 11). What is the kernel trick? why is it important? Medium. Retrieved November 3, 2022, from <https://medium.com/@zxr.nju/what-is-the-kernel-trick-why-is-it-important98a98db0961d>.
- [6] K.Sudhakar, and Dr. M. Manimekalai, January 2014, "Study of Heart Disease Prediction using Data Mining", International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 4, Issue 1,pp. 1157- 1160.
- [7] Kamal Kant, and Dr. Kanwal Garg,2014, "Review of Heart Disease Prediction using Data Mining Classifications",International Journal for Scientific Research & Development(IJSRD), Vol. 2, Issue 04, ISSN (online): 2321- 0613, pp. 109-111.
- [8] Vispute N. J. Sahu D. K. Rajput A. (2015, December). A Survey on naive Bayes Algorithm for Diabetes Data Set Problems.International Journal for Research in Applied Science and Engineering Technology, 3(12).
- [9] Vrushali Balpande Research on diabetes prediction method based on machine learning. J Phys Conf Ser. 2020;1684:1–6. [Google Scholar]
- [10] Le TM, Vo TM, Pham TN, Dao SV. A novel wrapper-based feature selection for early diabetes prediction enhanced with a metaheuristic. IEEE Access. 2020;9:7869–84. [Google Scholar]
- [11] Julius AO, Ayokunle AO, Ibrahim FO. Early diabetic risk prediction using machine learning classification techniques. Available from:<https://ijrsr.com/early-diabetic-risk-prediction-using-machine-learning-classification-techniques> . [Google Scholar]
- [12] Shafi S, Ansari GA. Early prediction of diabetes disease &classification of algorithms using machine learning approach. In Proceedings of the International Conference on Smart Data Intelligence (ICSMDI 2021) Available from:SSRN 3852590 (2021). [Google Scholar]
- [13] Khanam JJ, Foo SY. A comparison of machine learning algorithms for diabetes prediction. ICT Express. 2021;7:432–9. [Google Scholar]
- [14] Sisodia D, Sisodia DS. Prediction of diabetes using classification algorithms. Procedia Comput Sci. 2018;132:1578–85. [Google Scholar]
- [15] Norris SL, Lau J, Smith SJ, Schmid CH, Engelgau MM. Self-management education for adults with type 2 diabetes:A meta-analysis of the effect on glycemic control. Diabetes Care. 2002;25:1159–71. [PubMed] [Google Scholar]
- [16] Shaw JE, Sicree RA, Zimmet PZ. Global estimates of the prevalence of diabetes for 2010 and 2030. Diabetes Res Clin Pract. 2010;87:4–14. [PubMed] [Google Scholar]
- [17] Anjana RM, Pradeepa R, Deepa M, Datta M, Sudha V, Unnikrishnan R, et al. Prevalence of diabetes and prediabetes (impaired fasting glucose and/or impaired glucose tolerance) in urban and rural India:Phase I results of the Indian Council of Medical Research India Diabetes (ICMRINDIAB) study. Diabetologia. 2011;54:3022–7. [PubMed] [Google Scholar]
- [18] Y. Huang, Y. Chai, Y. Liu, J. Shen, Architecture of next-generation E-commerce platform. Tsinghua Sci. Technol. 24(1), 18–29 (2019).
- [19] Wagai GA, Romshoo GJ. Adiposity contributes to poor glycemic control in people with diabetes mellitus, a randomized case study, in South Kashmir, India. J Family Med Prim Care. 2020:4623–6. [PMC free article] [PubMed] [Google Scholar]
- [20] Mujumdar A, Vaidehi V. Diabetes prediction using machine learning algorithms. Procedia Comput Sci. 2019;165:292–9. [Google Scholar]
- [21] Diabetes mellitus affected patients classification and diagnosis through machine learning techniques. Procedia Comput Sci. 2017;112:2519–28. [Google Scholar]
- [22] Rawat V, Suryakant S. A classification system for diabetic patients with machine learning techniques. Int J Math Eng Manag Sci. 2019;4:729–44. [Google Scholar]
- [23] Perveen S, Shahbaz M, Guergachi A, Keshavjee K. Performance analysis of data mining classification techniques to predict diabetes. Procedia Comput Sci. 2016;82:115–21. [Google Scholar]





10.22214/IJRASET



45.98



IMPACT FACTOR:  
7.129



IMPACT FACTOR:  
7.429



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24\*7 Support on Whatsapp)